

1 CLAIMS

2 Having thus described our invention, what we claim as new and  
3 desire to secure by Letters Patent is as follows:

4 1. An information processing method comprising:

5 providing an annotation for multiple page files, including  
6 the steps of:

7 obtaining a plurality of page files from a web site;

8 generating a group of said page files, page layout  
9 structures of which are at least similar;

10 providing a first annotation for an arbitrary page file  
11 in said group; and

12 correlating said first annotation with at least a part  
13 of other page files of said group.

14 2. The information processing method according to claim 1,  
15 wherein said step of generating said group includes the steps  
16 of:

17 analyzing said page files to introduce structural  
18 descriptive forms for said page layout structures and  
19 characteristic values for said structural descriptive forms;

20 employing said structural descriptive forms and said  
21 characteristic values to calculate an inter-page distance  
22 representing a similarity of said page files; and

23 grouping said page files, of which said inter-page  
24 distance is equal to or smaller than a predetermined value.

25 3. The information processing method according to claim 2,

1 wherein said structural descriptive forms are layout tags  
2 employing a style for designating a location on a page for  
3 representing tags that are correlated with said page layout  
4 structures included in said page files; and wherein said  
5 characteristic values are attributes of said layout tags and  
6 values of said attributes.

7 4. The information processing method according to claim 2,  
8 wherein said inter-page distance is obtained by calculating a  
9 sum of the values obtained by weighting said characteristic  
10 value and said structural descriptive form that is included  
11 in common with said multiple page files.

12

13 5. The information processing method according to claim 1,  
14 wherein said step of correlating said first annotation with  
15 said other page files in said group includes the steps of:

16 determining whether said first annotation should be  
17 applied for the page files of said group;

18 adding a second annotation, when the determination is  
19 false, for an arbitrary page file of a page group consisting  
20 of page files with which said first annotation is not  
21 correlated;

22 correlating said second annotation with at least a part  
23 of other page files of said page group; and

24 correcting a calculation expression for said inter-page  
25 distance, so that, at said step of generating a group, said  
26 page file with which said first annotation is correlated and  
27 said page files that are correlated with said second  
28 annotation do not fall in the same group.

1   6. The information processing method according to claim 5,  
2 wherein said inter-page distance is calculated by using the  
3 sum of values obtained by weighting said characteristic value  
4 and said structural descriptive form that is included in  
5 common with said multiple page files; and wherein said  
6 calculation expression for said inter-page distance from a  
7 group of steps corrected by performing at least one step from  
8 a group of steps including:

9       an operation for increasing said weighting of said  
10 structural descriptive form and said characteristic value,  
11 for said structural descriptive form and said characteristic  
12 value that are different between said page file correlated  
13 with said first annotation and said page file correlated with  
14 said second annotation, and

15       an operation for reducing said weighting of said  
16 structural descriptive form and said characteristic value,  
17 for said structural descriptive form and said characteristic  
18 value that are in common with said page file correlated with  
19 said first annotation and said page file correlated with said  
20 second annotation.

21   7. The information processing method according to claim 2,  
22 further comprising the steps of:

23       introducing a representative structural descriptive form  
24 that represents said groups and a representative  
25 characteristic value for said representative structural  
26 descriptive form;

27       employing said representative structural descriptive  
28 form and said representative characteristic value to  
29 calculate an inter-group distance that delineates the

1 similarity between said groups;

2 grouping said page files that are included in said

3 groups, said inter-group distance of which is equal to or

4 smaller than a predetermined value, and generating a common

5 group;

6 adding an annotation to a common area wherein part of

7 the page layout structure of an arbitrary file, included in

8 common for the members of said common group, is the same as

9 or similar to at least a part of the page layout structure of

10 a different page file; and

11 correlating said annotation with said common area

12 provided for said different page file included, in common,

13 for said common group.

14 8. The information processing method according to claim 7,

15 wherein said representative structural descriptive forms are

16 layout tags employing a style for designating the location on

17 a page for representing tags correlated with said page layout

18 structures of said page files; and wherein said

19 representative characteristic values are attributes of said

20 layout tags and values of said attributes.

21

22 9. The information processing method according to claim 7,

23 wherein said inter-group distance is calculated by using the

24 sum of the values obtained by weighting said representative

25 characteristic value and said representative structural

26 descriptive form that is included in common with said

27 multiple groups.

28 10. The information processing method according to claim 7,

1 wherein said step of correlating said first annotation with  
2 said common area provided for said different page file  
3 includes the steps of:

4 determining whether said first annotation should be  
5 applied for said common area provided for the page files of  
6 said common group;

7 adding a second annotation, when the determination is  
8 false, to the common area of an arbitrary page file of a page  
9 group consisting of page files including said common area

10 with which said first annotation is not correlated;

11 correlating said second annotation with 'Yes' part of  
12 the common areas of other page files of said page group; and

13 correcting a calculation expression for said inter-group  
14 distance, so that, at said step of generating a common group,  
15 said page file including said common area correlated with  
16 said first annotation and said page files including said  
17 common areas correlated with said second annotation do not  
18 fall in the same common group.

19 11. An information processing system, for providing an  
20 annotation for multiple page files, comprising:

21 means for obtaining page files from a web site;

22 means for generating a group of said page files, page  
23 layout structures of which are the same or similar;

24 means for providing a first annotation for an arbitrary  
25 page file in said group; and

26 means for correlating said first annotation with 'Yes' a  
27 part of other page files of said group.

28 12. The information processing system according to claim

1 11, wherein said means for generating said group includes:  
2       means for analyzing said page files to introduce  
3 structural descriptive forms for said page layout structures  
4 and characteristic values for said structural descriptive  
5 forms;

6       means for employing said structural descriptive forms  
7 and said characteristic values to calculate an inter-page  
8 distance representing the similarity of said page files; and  
9       means for grouping said page files, of which said  
10 inter-page distance is equal to or smaller than a  
11 predetermined value.

12 13. The information processing system according to claim  
13 12, wherein said structural descriptive forms are layout tags  
14 employing a style for designating the location on a page for  
15 representing tags correlated with said page layout structures  
16 of said page files; and wherein said characteristic values  
17 are attributes of said layout tags and values of said  
18 attributes.

19 14. The information processing system according to claim  
20 12, wherein said inter-page distance is calculated by using  
21 the sum of the values obtained by weighting said  
22 characteristic value and said structural descriptive form  
23 that is included in common with said multiple page files.

24 15. The information processing system according to claim  
25 12, wherein said means for correlating said first annotation  
26 with said other page files in said group includes:  
27       means for determining whether said first annotation

1 should be applied for the page files of said group;  
2 means for adding a second annotation, when the  
3 determination is false, for an arbitrary page file of a page  
4 group consisting of page files with which said first  
5 annotation is not correlated;  
6 means for correlating said second annotation with 'Yes'  
7 part of other page files of said page group; and  
8 means for correcting a calculation expression for said  
9 inter-page distance, so that, at said step of generating a  
10 group, said page file correlated with said first annotation  
11 and said page files correlated with said second annotation do  
12 not fall in the same group.

13 16. The information processing system according to claim  
14 15, wherein said inter-page distance is calculated by using  
15 the sum of values obtained by weighting said characteristic  
16 value and said structural descriptive form that is included  
17 in common with said multiple page files; and wherein said  
18 calculation expression for said inter-page distance is  
19 corrected by performing at least one step from a group of  
20 steps including:

21 an operation for increasing said weighting of said  
22 structural descriptive form and said characteristic value,  
23 for said structural descriptive form and said characteristic  
24 value that are different between said page file correlated  
25 with said first annotation and said page file correlated with  
26 said second annotation, and  
27 an operation for reducing said weighting of said  
28 structural descriptive form and said characteristic value,  
29 for said structural descriptive form and said characteristic

1 value that are in common with said page file correlated with  
2 said first annotation and said page file correlated with said  
3 second annotation.

4 17. The information processing system according to claim  
5 12, further comprising:

6 means for introducing a representative structural  
7 descriptive form that represents said groups and a  
8 representative characteristic value for said representative  
9 structural descriptive form;

10 means for employing said representative structural  
11 descriptive form and said representative characteristic value  
12 to calculate an inter-group distance that delineates the  
13 similarity between said groups;

14 means for grouping said page files that are included in  
15 said groups, said inter-group distance of which is equal to  
16 or smaller than a predetermined value, and generating a  
17 common group;

18 means for adding an annotation to a common area wherein  
19 part of the page layout structure of an arbitrary file,  
20 included in common for the members of said common group, is  
21 the same as or similar to at least a part of the page layout  
22 structure of a different page file; and

23 means for correlating said annotation with said common  
24 area provided for said different page file included in common  
25 for said common group.

26 18. The information processing system according to claim  
27 17, wherein said representative structural descriptive forms  
28 are layout tags employing a style for designating the

1 location on a page for representing tags correlated with said  
2 page layout structures of said page files; and wherein said  
3 representative characteristic values are attributes of said  
4 layout tags and values of said attributes.

5 19. The information processing system according to claim  
6 17, wherein said inter-group distance is calculated by using  
7 the sum of the values obtained by weighting said  
8 representative characteristic value and said representative  
9 structural descriptive form that is included in common with  
10 said multiple groups.

11 20. The information processing system according to claim  
12 17, wherein said means for correlating said first annotation  
13 with said common area provided for said different page file  
14 includes:

15       means for determining whether said first annotation  
16 should be applied for said common area provided for the page  
17 files of said common group;

18       means for adding a second annotation, when the  
19 determination is false, to the common area of an arbitrary  
20 page file of a page group consisting of page files including  
21 said common area with which said first annotation is not  
22 correlated;

23       means for correlating said second annotation with 'Yes'  
24 part of the common areas of other page files of said page  
25 group; and

26       means for correcting a calculation expression for said  
27 inter-group distance, so that, at said means for generating a  
28 common group, said page file including said common area

1 correlated with said first annotation and said page files  
2 including said common areas correlated with said second  
3 annotation do not fall in the same common group.

4 21. An article of manufacture comprising a computer usable  
5 medium having computer readable program code means embodied  
6 therein for causing annotation, the computer readable program  
7 code means in said article of manufacture comprising computer  
8 readable program code means for causing a computer to effect  
9 the steps of claim 1.

10 22. A program storage device readable by machine, tangibly  
11 embodying a program of instructions executable by the machine  
12 to perform method steps for annotation said method steps  
13 comprising the steps of claim 1.

14 23. A computer program product comprising a computer usable  
15 medium having computer readable program code means embodied  
16 therein for causing annotation the computer readable program  
17 code means in said computer program product comprising  
18 computer readable program code means for causing a computer  
19 to effect the functions of claim 11.